Whitman
SCHOOL of MANAGEMENT
SYRACUSE UNIVERSITY

MBC 638
DATA ANALYSIS AND DECISION MAKING

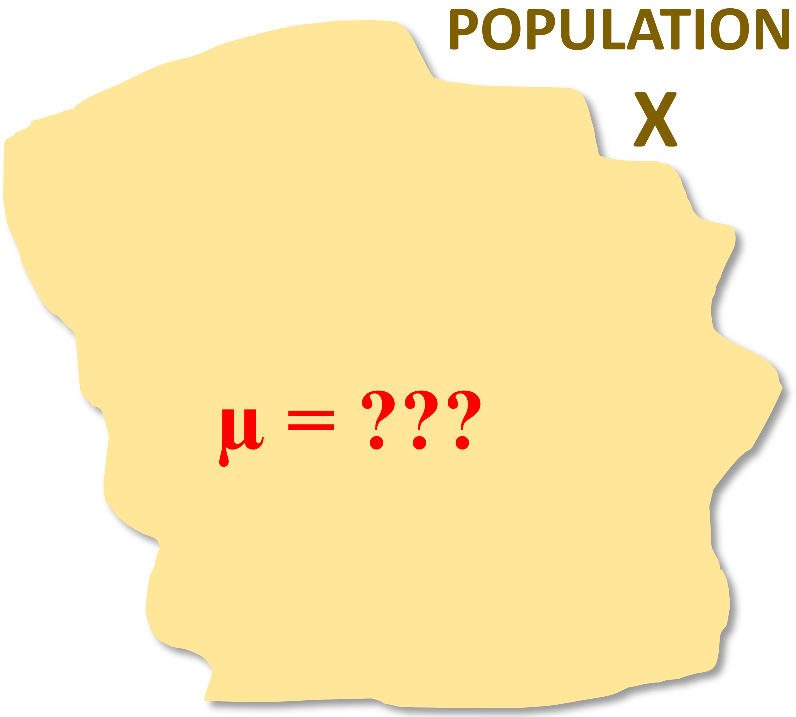
Anna Chernobai, Department of Finance

HYPOTHESIS TESTING for population mean
Chapter 9

Today:

Hypothesis testing for μ

We have a sample and we want to use the **sample mean $\bar{x}$** to estimate the unknown **population mean μ**.



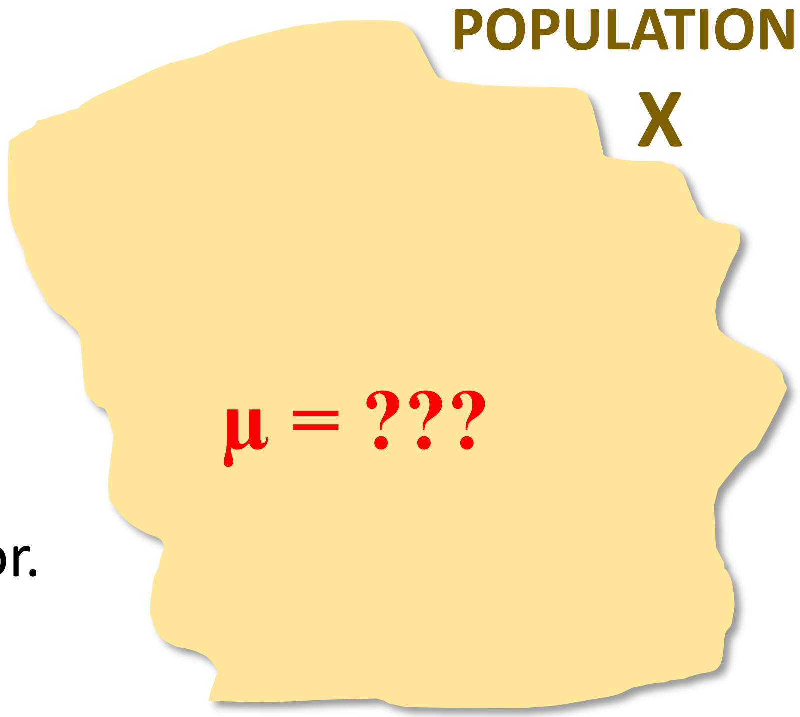
<u>Example</u>:

We are interested in the national average spending amount per person during Halloween.

X = Halloween spending per person

μ = average Halloween spending per person nationally

We have a sample and we want to use the **sample mean $\bar{x}$** to estimate the unknown **population mean μ**.



- In C.L.T., we saw that we can use sample mean as a **point estimate** to estimate the population mean, but with an error.

- In last topic (C.I.), we learnt about the **interval estimate** of the population mean.

  Specifically, we estimated an interval such that, with some high probability (e.g., 95%) we are certain that the true population mean μ lies within this interval.

- In this topic, we have some specific **hypothesis** about the population mean, and we need to determine whether this sample provides evidence whether this hypothesis is **supported** or **not supported**.

Confidence Intervals and Hypothesis Testing are very closely related.

Often, we don't care about the values that μ may take (confidence interval). All we care is whether some specific hypothesis about μ is supported or is not supported by our sample data.

- In last topic (C.I.), we learnt about the **interval estimate** of the population mean.

  Specifically, we estimated an interval such that, with some high probability (e.g., 95%) we are certain that the true population mean μ lies within this interval.

- In this topic, we have some specific **hypothesis** about the population mean, and we need to determine whether this sample provides evidence whether this hypothesis is **supported** or **not supported**.

Example

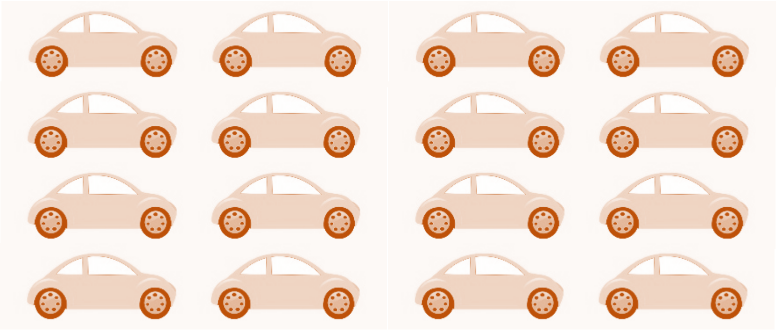
Example

A car dealer calculated that the company must average 4.8% or more profit on the sales of its new cars to cover its expenses. A random sample of 80 cars produced an average profit of 4.9% per car. The historical standard deviation is known to be 0.3%.

GOAL

μ = **4.8%** **or more**

average profit per car



Sample:

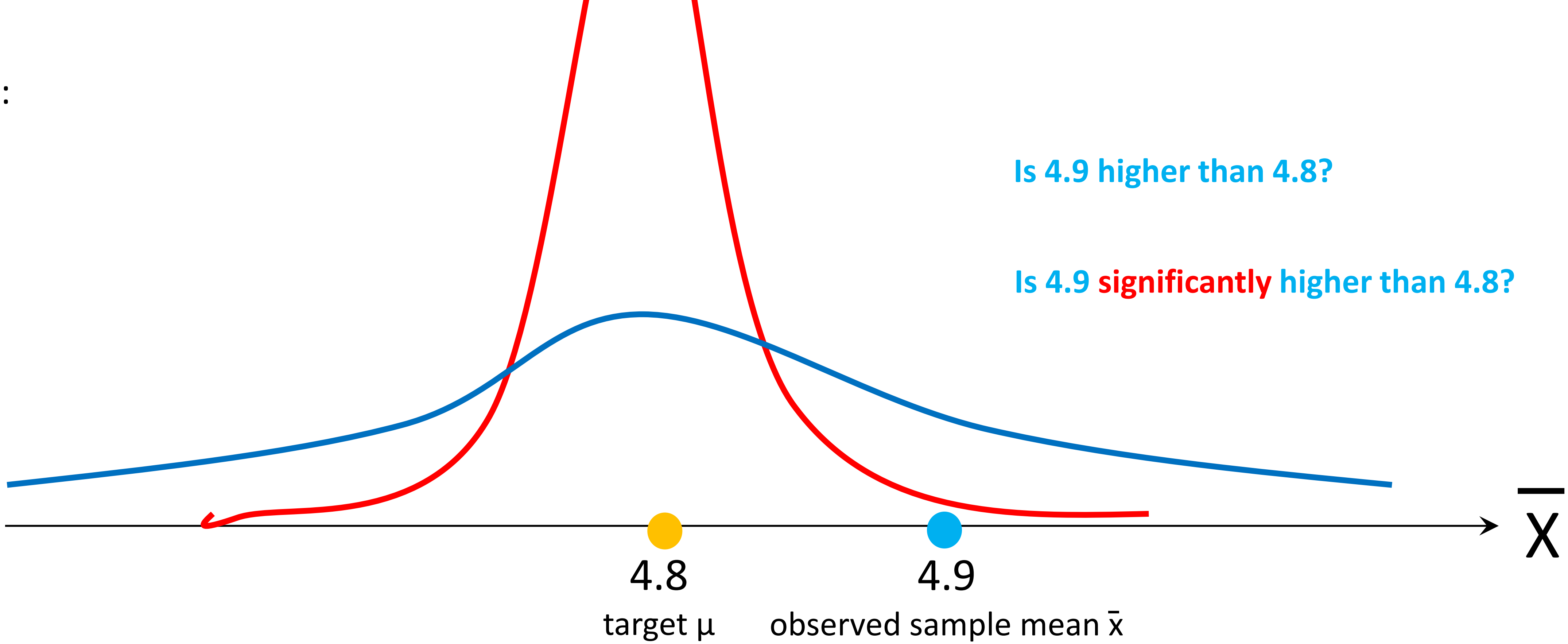
$\bar{x}$ = **4.9%**

<u>The car dealer observes</u>: observed sample mean > target mu

<u>Formulate hypothesis:</u> Does the observed sample provide sufficient evidence to indicate that the sales policy is achieving a **true mean** profit **exceeding 4.8%** per car?

Hypothesis: Does the observed sample provide sufficient evidence to indicate that the sales policy is achieving a **true mean** profit **exceeding 4.8%** per car?

Sample:
$\bar{x} = 4.9$



Main idea: If we manage to show that $\bar{x}=4.9$ is an **outlier** (unusually large) relative to 4.8, then our hypothesis is supported and we conclude that we have evidence that the **true** μ is > 4.8. If not, then our hypothesis is not supported.

Let’s solve this example using the formal **hypothesis testing** procedure

Statistical procedure:

Step 1: Summarize the hypotheses:

Null hypothesis:	H_0:	$\mu \leq 4.8$
Alternative hypothesis:*	H_A:	$\mu > 4.8$

In a sense, you are always testing 2 hypotheses in parallel, but only one of them will be supported.

The alternative hypothesis (H_A) summarizes what you want to test.

Statistical procedure:

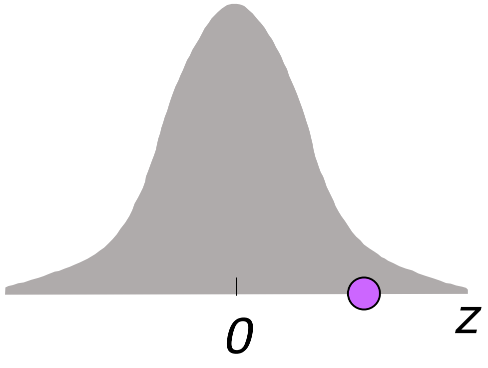
<u>Step 2:</u> Calculate the test statistic:

$$Test\ Statistic = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

It's the **Z score** of the observed x̄ assuming that μ_0 is true

$\mu_0 = 4.8$
$\bar{x} = 4.9$
$\sigma = 0.3$
$n = 80$

$$Test\ Statistic = \frac{4.9 - 4.8}{0.3 / \sqrt{80}} = 2.98$$



Statistical procedure:

<u>Step 3:</u> Calculate the ***p***-value:

$$P(z > Test\ Statistic)$$

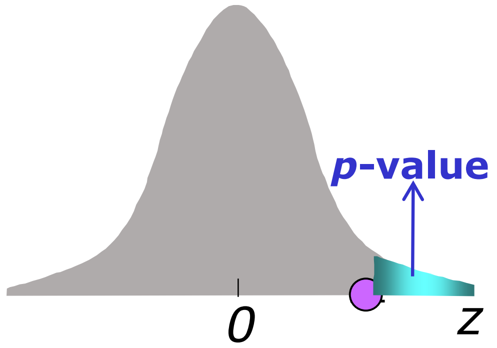
$$P\left(z > \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}\right)$$

$\mu_o = 4.8$
$\bar{x} = 4.9$
$\sigma = 0.3$
$n = 80$

$$P\left(z > \frac{4.9 - 4.8}{0.3/\sqrt{80}}\right)$$

$$= P(z > 2.98)$$

$$= 0.0014$$



Statistical procedure:

<u>Step 4:</u> Compare ***p***-value with **α** and state your conclusions:

α = the maximum error you are willing to tolerate (is given, e.g., 5%)

= Probability (H_0 is rejected while H_0 is actually true)

***p*-value** = the probability of seeing a random sample *at least as extreme as the observed sample*, assuming that the null hypothesis is true

= the measure of evidence against the null hypothesis H_0

Smaller *p*-values indicate more evidence against H_0

Decision rule: **Reject H_0 if *p*-value < *α***

REALITY

RESULTS OF HYPOTHESIS TESTING

	Ho is true	Ho is false
Ho is not rejected	CORRECT	Type II Error
Ho is rejected	$\boxed{\text{Type I Error}}$	CORRECT

$$\alpha = \text{Prob (Type I Error)}$$

Example: Putting it all together

A car dealer calculated that the company must average 4.8% profit on the sales of its new cars to cover its expenses. A random sample of 80 cars produced an average profit of 4.9% per car. The historical standard deviation is known to be 0.3%.

Based on the observed sample mean, does the data provide sufficient evidence to indicate that the sales policy is achieving a **true** mean profit **exceeding** 4.8% per car? Use α=1%.

- Step 1: H_0: $\mu \leq 4.8$
  H_A: $\mu > 4.8$
- Step 2: Compute test statistic: $Z = \dfrac{4.9 - 4.8}{0.3/\sqrt{80}} = 2.98$
- Step 3: Compute *p*-value: $pvalue = P\left(z > \dfrac{4.9 - 4.8}{0.3/\sqrt{80}}\right) = P(Z > 2.98) = 0.0014$
- Step 4: Compare *p*-value=0.0014 with α=0.01. *P*-value < α. **Reject H_0**. We have **sufficient evidence** to conclude that the company is indeed achieving policy of over 4.8% average profit per car. ☺

Rule: $\boxed{\text{Reject } H_0 \text{ if p-value} < \alpha}$

Left-tailed	Right-tailed	2-Tailed
Hypothesis Testing for μ (population mean)		
Step 1: Formulate hypotheses:		
H_0: $\mu \geq \mu_0$ H_A: $\mu < \mu_0$	H_0: $\mu \leq \mu_0$ H_A: $\mu > \mu_0$	H_0: $\mu = \mu_0$ H_A: $\mu \neq \mu_0$
Step 2: Compute test statistic:		
$Test\ Statistic = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$	$Test\ Statistic = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$	$Test\ Statistic = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$
Step 3: Compute p-value:		
$p\text{–}value = P(Z < Test\ Statistic)$	$p\text{–}value = P(Z > Test\ Statistic)$	$p\text{–}value = 2 \bullet P(Z > \lvert Test\ Statistic \rvert)$
(1) σ known: Use Standard Normal distribution (2) σ unknown, n≥30: Use *s* for standard deviation and *t*-distribution, d.f.=n-1 (or Normal approx.) (3) σ unknown, n<30: Use *s* for standard deviation and *t*-distribution, d.f.=n-1		
Step 4: Reject H_0 if p-value $< \alpha$		

Example

Annual per capita consumption of milk is 21.6 gallons (*Statistical Abstract of the United States* 2006).

A sample of 76 individuals from California showed an average annual consumption of 24.1 gallons with a standard deviation of 4.8.

a) Based on this observation, you believe milk consumption is higher in California than the national average and would like to test this hypothesis. Use α=0.05.

Example

Annual per capita consumption of milk is 21.6 gallons (*Statistical Abstract of the United States* 2006).

A sample of 76 individuals from California showed an average annual consumption of 24.1 gallons with a standard deviation of 4.8.

b) Based on this observation, you believe milk consumption is different in California from the national average and would like to test this hypothesis. Use $\alpha=0.05$.

c) Construct a 95% confidence interval for the average annual consumption of milk in California. Can you see how this interval reinforces your conclusion in part b)?

Today:

Hypothesis testing for μ